



开篇词 | 实时互动强势发展，如何快速入门音频技术？

2021-11-22 冯建元

《搞定音频技术》

[课程介绍 >](#)



讲述：冯建元

时长 10:10 大小 9.32M



你好，我是建元。桃李春风一杯酒，江湖夜雨十年灯。乘着算法的一叶扁舟在多个行业的江、河、湖、海中，我不知不觉已经遨游了将近十年。

一名“算法吟游诗人”的 10 年

如果用一句话来总结我的经历，我想我应该是算法吟游诗人，我会随时将灵感运用于我的工作中。而与音频领域的结缘，可以回溯到 2011 年。

那时候我还在国内读大学，去工厂实习时，我记得有位老师傅，他只需要用耳朵听一☆就能判断出设备是哪里出了故障。受此启发，我当时是利用人工神经网络和振动传感器，搭建了一套空气压缩机故障自动诊断系统。那时，我惊叹于傅立叶变换的时、频域变换的神奇，人工神经网络的反向传播机制以及算法作为交叉学科实践的魅力。

后来为了进一步学习，我去了美国芝加哥的 Illinois Institute of Technology，攻读了硕士和博士学位。在硕博期间，我逐渐发现各个领域的算法其实都是相通的。利用信号处理、控制理论、人工智能算法的融合，我陆续研发了一系列包括毛细流体力学建模、智能电网系统、人工智能胰岛系统在内的多个专业的应用，并在 IEEE、AIChE、Journal of Process Control 等一线期刊上发表了 20 多篇论文。

也正是因为这些专业的学习，为我后面的工作乃至创业都打下了一个很好的基础。

博士毕业后，我去了 University of Illinois at Chicago，担任特聘研究员。这段时间也正好是人工智能网络开始在音视频、自然语言处理等领域崭露头角的时候。那时我正好经手了**两个项目**：一个是基于自然语言处理的医学大数据的智能诊断，另一个是基于图像的食物营养成分的提取。我意识到传统算法可能在 AI 算法的加持下会迈入一个新的纪元，而且正好国内工业 5.0 正在如火如荼的改革进程中。于是，我便回国作为合伙人加入了一家做工业音视频智能化改造的创业公司。

创业期间，我在音频方面做了很多有意思的尝试。比如，利用多个麦克风阵列矩阵，实现园区的异常声音的预警系统。该系统可以根据麦克风阵列采集的音频信号自动判断园区中哪个区域、哪台设备的什么地方发生了故障，从而实现了全产线的无人化管理。

后续结合音视频和机器人，我还做了一套自动巡检机器人系统，让机器人像导游一样在区域内巡逻，帮助访客找到自己想去的地点。机器人集合的音频系统包括语音识别、对话交互和语音生成，几乎每个环节都是传统信号处理算法和 AI 算法的联合应用。现在类似的机器人，我想你在生活中也接触到了。

创业阶段让我完整地认识了企业的运作和很多不同行业涌现出的新需求、新玩法，这是件好事。但后来我觉得那段时间做的东西过于繁杂了，反而不容易把一件事情做到极致。我就开始思考，**什么样的行业能包含比较多的领域，又不至于失去趣味，同时又可以长期研究、不断精进呢？**

最终，结合之前的经历，我选择了音频，摆脱了创业的繁琐，加入声网——这家把“声音”放入名字的公司。后来在这里，我研发了一系列实时声音美化、实时空间音频渲染、实时变声、实时修音、实时语音变歌唱等算法，并先后带领团队完成并商用了 50 多种可实时使用的美声音效效果。

之后我又对实时音频链路进行了 AI 化改造，包括基于深度学习的降噪、音频编 / 解码、音频丢包补偿、回声消除、音源分离、啸叫检测、音乐检测等算法的实现与落地商用，并把这些算法和不同的业务场景相结合提高实时音频质量。同时，我也是国内第一个基于深度学习的语音编解 / 码器 Silver 的发明人。

那听了我这这么多的故事，我想我们有必要重新认识一下了。你好，我是声网的音频算法专家，你叫我建元就好。那么提到音频技术，我为什么鼓励你深度学习它呢？

为什么说音频是实时互动开发中的必备武器？

从整个大环境看，音频已然是实时互动开发中的必备武器。

在新冠疫情的催化下，线上互动已经成为人们工作和生活中的必要交流方式。我记得 2020 年疫情爆发的时候，大部分的企业都在远程办公，学生们也在接受在线教育，这催生了很多音频技术的进一步完善。比如说，在这些场景中，延迟必须控制在毫秒级别内，实时互动技术成为决定整体体验的关键，从某种层面来说也是考验着大家的底层实时音视频（RTC）技术。

而在实时互动中，音频承载的信息密度又是巨大的。语言、音乐都需要音频作为载体进行传播。而人耳对声音的感知又是如此敏感，一个音频采样点的错误，人耳就能明显地感知到。那么像噪声、回声、网络抖动等问题，就更需要专业的处理了。随着 AI 技术的介入，例如基于 AI 降噪、回声消除、音频编解码技术，互动音频的可靠性已经得到了更多的保障。**学习它，恰逢其时。**

而随着大家对线上体验的增加，人们对互动体验中的音频体验要求也在提升。人们逐渐从听得到、听得清，向好听、音质还原度高等方向发展。那么美声、音效技术就可以使你的声音更好听；高音质实时编解码器搭配优秀的抗弱网策略可以让音乐会、合唱等场景从线下搬到线上；空间音频技术能让你在虚拟世界中实现“声临其境”……可以说，**音频互动技术就像空气和水一样无处不在，且充满了无限的可能。**

近些年，除了短视频、直播等传统方向会用到音频的处理，我们看到在新体验、新玩法的领域中音频也成了一件必备武器。例如，Facebook 改名 Meta 进军元宇宙，音频播客 Clubhouse 等纯音频社交方式的流行，以及 TWS 耳机支持了空间音频渲染和主动降噪等等。

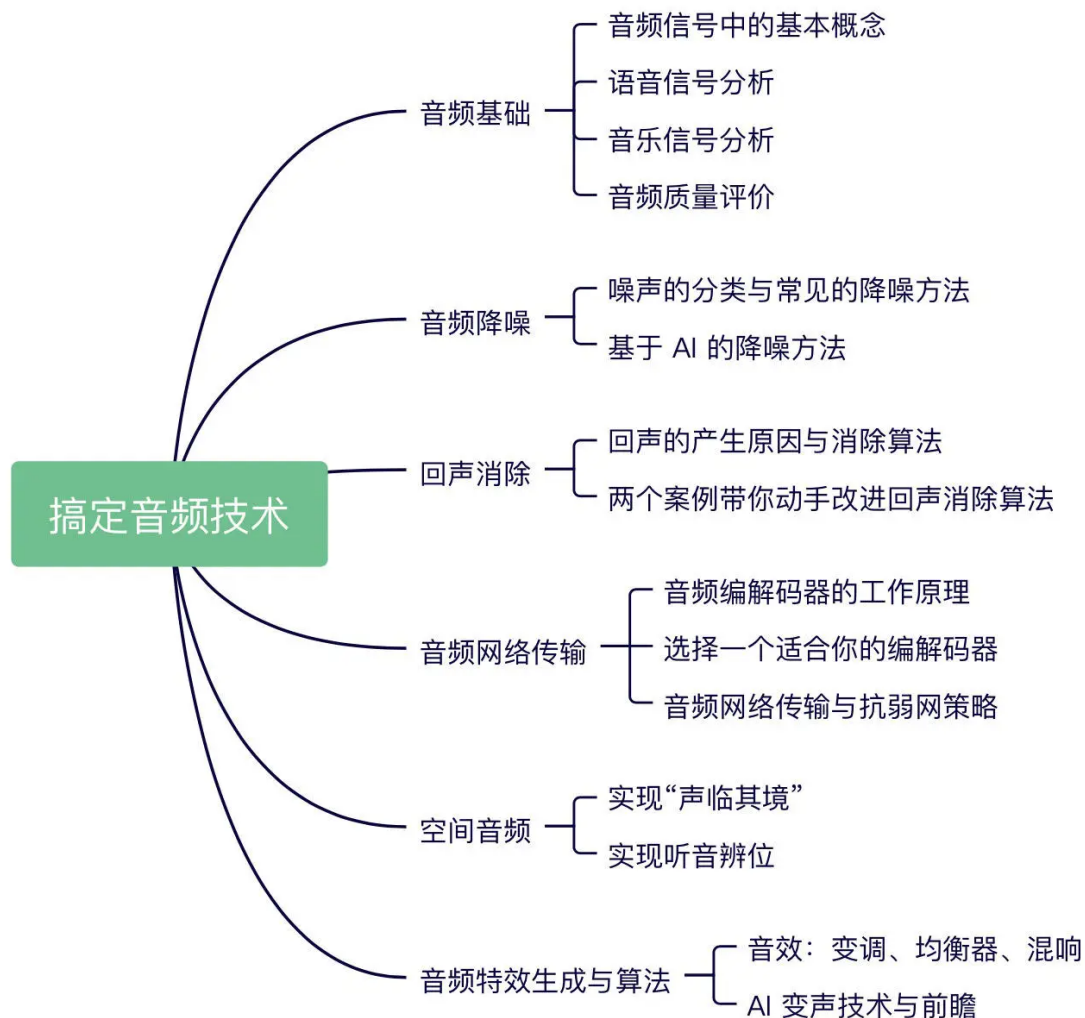
而在未来，可预测的是，人工智能领域在音频技术上会有进一步的突破。大型线上互动、娱乐、会议、教育等行业万象都会朝着音视频实时互联互通的方向转变。市场对于专业音频技术专家、音频工程师的需求缺口也在日渐增大。音频领域和计算机、信号处理、声学、音乐、语言学等都有交集，**你过往的经历都可能派上用处。**

课程设计

音频是一个跨专业领域的有趣学科。而兴趣驱动，永远是最强的动力。所以，无论你是否是专业的音频工程师，有没有强大的技术背景和积累，都不要紧。只要你有想法转行音频领域，甚至只要你有兴趣学习，这门课都能给你创造新的机会。

因为我们每天都在说话、都在听各种各样的声音。我们平时发音有什么规律？耳朵对不同频段的心理感知有什么区别？乐器又是怎么发音的？我们如何判断音频的质量？以及“好声音”是怎么修出来的.....这些知识不但可以让你用于职业工作，也可以加强你对生活中声音的感知，从而知道应该听什么、怎么听，甚至怎么去让声音发生变化，为你打开“百万调音师”的潘多拉宝盒。

那说了这么多，这门课的具体内容到底是怎样的呢？**我分为了 6 个模块，设计思路是这样的。**



首先，知识基础决定上层应用。所以我们的课程会先从音频的基础知识着手，带你了解音频是如何产生的，我们的听觉是怎么感知语音和音乐的，以及我们是如何评价音频质量的好坏的。

其次，有了基础知识以后，我会带你看看实时音频处理中所有的常见模块，包括：音频降噪、回声消除、音频编解码器的原理和选择，以及音频抗弱网策略等。有了这些知识你将会对整个音频链路有一个完整的认识。其中，我也会分享很多实践案例，帮助你利用算法解决各种可能遇到的音频问题。这里你既能学到很多音频信号处理的算法原理，也能从诸如 AI 降噪算法的设计中了解到音频处理算法中的前沿方向。

最后，我将带你看看 VR/AR 音频中常见的空间音频是如何渲染的，揭秘音效制作的基本原理，以及通过探究 AI 变声算法原理，带你看看 AI 究竟在音频领域还有哪些有意思的应用。

总结来说，通过这门课程的学习，你能知晓音频处理的全套链路和音频发展的脉络、方向。这里也给你一个学习指导，你可以在学习过程中重点关注。

对于基础，我会尽量讲解得通俗易懂，但每一讲都需要你加强记忆，后面还会用到；

对于音频处理，我会抽丝剥茧地带你看看技术演进的过程，对比不同技术的优缺点，以便于你今后的应用；

而对于音频算法和应用，这些都是不断演化的，我会带你了解目前音频前沿领域的新玩法和新探索，让你能从音频领域中一个更靠前的起跑线开始奔跑；

对于你想进一步了解的内容，我也会通过大量的相关文献链接，以及课后思考题的形式帮助你充分理解。

最后，我想说音频技术作为一个很贴近生活的专业领域，实用性和趣味性都很出众，且这方面的人才一直十分稀缺。从 AI 技术的发展和虚拟现实技术的进步来看，音频都是一个充满朝阳的方向。如果你想转型或者拓展你的技术能力，加入我，让我们一起推开音频这扇新世界的大门。

也欢迎你在留言区中与我交流，不妨说说你对这门课程的期待，或者你在学习音频技术时的经历和痛点，我会更加有针对性地为你的答疑解惑！

分享给需要的人，Ta 订阅后你可得 **20 元现金奖励**

 生成海报并分享

 赞 2

 提建议

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

下一篇 01 | 声音是如何保存成数字信号的？

精选留言 (6)

 写留言



Fan

2021-11-22

什么开发语言的呢

展开 ∨

作者回复: 音频的开发语言取决于你所在的行业场景和工作内容。工程方面比如硬件开发可能更多的和C打交道。为了保证实时性和性能的移动端APP或者SDK那么C++用的比较多。如果是做算法研发那么python和Matlab会是首选。



3



怕什么，抱紧我

2021-11-27

音乐系毕业，java路过，需要什么基础呢！就懂点乐理，音高，音频，音色，五度相生律，十二平均律相关的！

展开 ∨

作者回复: 音乐毕业入行的也有不少，音频算法相关的不少和音乐知识有联系。这门课还是比较基础的只需要少许信号处理的基础知识，结合自己的专业可以在修音、空间音频、音乐编解码等方向都会有很多的新的发现。



1



哈珀朋友

2021-12-03

希望声网能为中国音视频的技术发展做出突出贡献

展开 ∨



自由

2021-11-30

冯老师，你好，我的主语言是 go，c、python 我用的比较少，请问订阅后能结合专栏知识，实现什么功能呢？

展开 ∨

作者回复: 想往音频分析算法方面转的话建议可以把python用起来，这样可以有很多现成的算法开源库可以使用。至于最后使用go还是C还是java做成工程化的产品功能就应人和产品而异了。



**pencilCool**

2021-11-28

我自己从小就喜欢音乐，会玩一些乐器，在学校学的也是信号处理，现在主要搞ios 开发，这两年希望转到音频技术的研发。老师这们课必须第一时间跟上呀。

作者回复: 那其实基础不错，驾轻就熟~

**刘宏鑫**

2021-11-24

因为当前公司是做声纹技术的，日常开发任务也包括很多音视频的处理工作，和FFmpeg打交道比较多，最近想深入学习音视频的相关知识，感觉老师讲得针不错，不过我是来催更的哈哈哈

编辑回复: 加油！共同学习，每周一三五更新哟

